MUTUAL INFORMATION APPROACH TO BLIND SEPARATION-DECONVOLUTION

Dinh-Tuan Pham

Laboratory of Modeling and Computation, INPG/UJF/CNRS
B.P. 53X, 38041 Grenoble cedex 9, France
phone:+33 4 76 51 44 23, fax: +33 4 76 63 12 63, email:Dinh-Tuan.Pham@imag.fr
web: www-lmc.imag.fr/lmc-sms/Dinh-Tuan.Pham

ABSTRACT

In this paper, we propose an approach to the blind separation-deconvolution problem, based on the mutual information criterion. Formulas for the quasi Newton algorithm are provided. More interesting results have been obtained in the pure deconvolution case. By a clever parameterization of the deconvoluting filter, the quasi Newton algorithm also become particularly simple. Simulation results showing the good performance of this algorithm are provided.

1. INTRODUCTION

This paper explores the use of the mutual information for the blind separation and deconvolution of convolutive mixtures. Specifically, K observed sequences $\{X_1(t)\},\dots\{X_K(t)\}$ are related to K source sequences[1] $\{S_1(t)\},\dots\{S_K(t)\}$ via linear convolutions

$$\mathbf{X}(t) = \sum_{u=-\infty}^{\infty} \mathbf{A}(u)\mathbf{S}(t-u), \qquad (1)$$

where $\mathbf{X}(t)$ and $\mathbf{S}(t)$ denote the vectors with components $X_1(t),\dots,X_K(t)$ and $S_1(t),\dots,S_K(t)$, respectively, and $\{\mathbf{A}(u)\}$ is a sequence of square matrices. The goal is to recover the sources from the observations and naturally this is done through the use of a sepaarating filter

$$\mathbf{Y}(t) = \sum_{u=-\infty}^{\infty} \mathbf{B}(u)\mathbf{X}(t-u), \qquad (2)$$

where $\{\mathbf{B}(u)\}$ is a sequence of matrices to be determined. The components $Y_k(t)$ of the vector $\mathbf{Y}(t)$ represent the recovered sources and since no specific knowledge of the sources is available (in a blind context), the idea is to exploit their independence assumption and thus determine $\{\mathbf{B}(u)\}$ such that the sequences $\{Y_1(t)\},\dots,\{Y_K(t)\}$ are as (mutually) independent as it is possible. This principle can only seperate the sources up to a filtering, since replacing each of the sequences $\{Y_1(t)\},\dots,\{Y_K(t)\}$ by a corresponding filtered version would not destroy their independence. However, in the deconvolution problem, the sources are temporally independent and one may require that the sequences $\{Y_1(t)\},\dots,\{Y_K(t)\}$ are also as temporally independent as it is possible (beside being mutually independent), We call this problem the *blind separation-deconvolution* problem.

A well known popular measure of dependence is the mutual information criterion, which provide efficient separation in the case of instantaneous mixtures [1]. Here, we shall adopt as criterion the average mutual information

$$\lim_{L\to\infty} \frac{1}{L} I[Y_k(1),\dots,Y_k(L), k=1,\dots,K],$$

$Y_k(t)$ are the components of $\mathbf{Y}(t)$ defined in (2) and $I(\cdots)$ denotes the mutual information between the indicated random variables. But it is well known that the mutual information can be expressed in term of entropy: $I(Z_1,\dots,Z_K) = \sum_{j=1}^K H(Z_j) - H(Z_1,\dots,Z_K)$ where $Z_1,\dots,Z_K$ are random vectors and $H(\cdot)$ denotes the entropy (or joint entropy when appropriate)[2]. Further, $H[\mathbf{Y}(1),\dots,\mathbf{Y}(L)]/L$ converges as $L\to\infty$ to entropy rate of the process $\{\mathbf{Y}(t)\}$ [2], denoted by $H[\mathbf{Y}(\cdot)]$. Therefore the above criterion can be written as $\sum_{k=1}^K H[Y_k(\cdot)] - H[\mathbf{Y}(\cdot)]$. But [3]

$$H[\mathbf{Y}(\cdot)] = H[\mathbf{X}(\cdot)] + \int_0^{2\pi} \log|\det \mathbf{B}(\omega)| \frac{d\omega}{2\pi},$$

where $\mathbf{B}(\omega) = \sum_{u=-\infty}^{\infty} \mathbf{B}(u)e^{-i\omega u}$. Note that for simplicity, we have used the same symbol $\mathbf{B}$ both in $\{\mathbf{B}(u)\}$ to denote the filter impulse response and in $\mathbf{B}(\omega)$ to denote its frequency response, the variable u (roman letter) and ω (Greek letter) help to avoid the confusion. Finally, the criterion equals, up to an additive constant

$$C(\mathbf{B}) = \sum_{k=1}^{K} H(Y_k) - \int_0^{2\pi} \log|\det \mathbf{B}(e^{i\omega})| \frac{d\omega}{2\pi}. \qquad (3)$$

Such criterion have been used in [4] (and in [5] but without the determinant term) in the pure deconvolution case ($K=1$).

2. THE EMPIRICAL CRITERION AND ESTIMATING EQUATIONS

In practice, one has to replace the criterion $C(\mathbf{B})$ with an estimator and this is naturally done by replacing $H(Y_k)$ by an entropy estimator $\hat{H}(Y_k)$, such as the one proposed in [6]. As only a finite length record, $\mathbf{X}(0),\dots,\mathbf{X}(T-1)$, say, is observed while the definition (2) may require the knowledge of the entire sequence $\{\mathbf{X}(t)\}$, we shall extend periodically the observation and thus redefine

$$\mathbf{Y}(t) = \sum_{u=-\infty}^{\infty} \mathbf{B}(u)\mathbf{X}(t-u \bmod T). \qquad (4)$$

[1] we restrict ourselves the case where there are a same number of sources as sensors, for simplicity

[2] The entropy of a random vector (or variable) $\mathbf{Z}$ with density $p_\mathbf{Z}$ is defined as $-\int p_\mathbf{Z}(\mathbf{z})\log p_\mathbf{Z}(\mathbf{z})d\mathbf{z}$, the joint entropy of several random vectors (or variables) is the entropy of the vector obtained by stacking their components

This yields a *periodic sequence of period T* and the entropy estimator $\hat{H}(Y_k)$ is computed in terms of $Y_k(0),\ldots,Y_k(T-1)$. We take as empirical criterion

$$\hat{C}(\mathbf{B}) = \sum_{k=1}^{K} \hat{H}(Y_k) - \frac{1}{T}\sum_{n=0}^{T-1} \log\det \mathbf{B}\Big(\frac{2\pi n}{T}\Big). \tag{5}$$

We now assume that the matrix sequence $\{\mathbf{B}(u)\}$ is parameterized by some parameter θ and we are interested in the gradient and the criterion $\hat{C}(\mathbf{B})$.

It was shown in [3] that $\lim_{\varepsilon\to 0}[H(Y+\varepsilon Z)-H(Y)]/\varepsilon = \mathrm{E}[\psi_Y(Y)Z]$ where ψ_Y is the negative of the logarithmic derivative of the density of Y, and is known as the score function of Y. Thus, following [6] we estimate ψ_{Y_k} through the partial derivative of $\hat{H}(Y_k)$: $\hat{\psi}_k[Y_k(t)] = T\partial\hat{H}(Y_k)/\partial Y_k(t)$. Then denoting by $\hat{\psi}[\mathbf{Y}(t)]$ the vector with components $\hat{\psi}_1[Y_1(t)],\ldots,\hat{\psi}_K[Y_K(t)]$, the gradient of $\sum_{k=1}^{K}\hat{H}(Y_k)$ equals $\mathrm{tr}\{T^{-1}\sum_{t=0}^{T-1}[\partial\mathbf{Y}(t)\partial\theta_\mu]\hat{\psi}[\mathbf{Y}(t)]^{\mathrm{T}}\}$ where

$$\frac{\partial\mathbf{Y}(t)}{\partial\theta_\mu} = \sum_{u=-\infty}^{\infty}\frac{\partial\mathbf{B}(u)}{\partial\theta_\mu}\mathbf{X}(t-u \bmod T). \tag{6}$$

and $^{\mathrm{T}}$ denotes the transpose. It follows that the gradient of the criterion (5) equals

$$\sum_u \mathrm{tr}\Big\{\frac{\partial\mathbf{B}(u)}{\partial\theta_\mu}\frac{1}{T}\sum_{t=0}^{T-1}\mathbf{X}(t-u \bmod T)\hat{\psi}^{\mathrm{T}}[\mathbf{Y}(t)]\Big\}$$
$$-\frac{1}{T}\sum_{n=0}^{T-1}\mathrm{tr}\Big\{\frac{\partial\mathbf{B}(2\pi n/T)}{\partial\theta_\mu}\mathbf{B}\Big(\frac{2\pi n}{T}\Big)^{-1}\Big\} \tag{7}$$

Alternatively, by the (discrete) Parseval equality

$$\sum_{t=0}^{T-1}\frac{\partial\mathbf{Y}(t)}{\partial\theta_\mu}\hat{\psi}[\mathbf{Y}(t)]^{\mathrm{T}} = \frac{1}{T}\sum_{n=0}^{T-1}\mathbf{d}_{\partial\mathbf{Y}/\partial\theta_\mu}\Big(\frac{2\pi n}{T}\Big)\mathbf{d}^{\mathrm{T}}_{\hat{\psi}(\mathbf{Y})}\Big(\frac{-2\pi n}{T}\Big)$$

where $\mathbf{d}_{\partial\mathbf{Y}/\partial\mu}(2\pi n/T) = \sum_{t=0}^{T-1}[\partial\mathbf{Y}(t)/\partial\theta_\mu]e^{-i2\pi nt/T}$, $n = 0,\ldots,T-1$, is the discrete Fourier transform of the sequence $\partial\mathbf{Y}(t)/\partial\theta_\mu,\ldots,\partial\mathbf{Y}(T-1)/\partial\theta_\mu$ and $\mathbf{d}_{\hat{\psi}(\mathbf{Y})}$ is defined similarly. Therefore, noting that by (6), $\mathbf{d}_{\partial\mathbf{Y}/\partial\theta_\mu}(2\pi n/T) = [\partial\mathbf{B}(2\pi n/T)/\partial\theta_\mu]\mathbf{d}_{\mathbf{X}}(2\pi n/T)$ and by (4), $\mathbf{d}_{\mathbf{X}}(2\pi n/T) = \mathbf{B}(2\pi n/T)^{-1}\mathbf{d}_{\mathbf{Y}}(2\pi n/T)$, one gets an alternative expression for the gradient:

$$\frac{1}{T}\sum_{n=0}^{T-1}\mathrm{tr}\Big\{\frac{\partial\mathbf{B}(2\pi n/T)}{\partial\theta_\mu}\mathbf{B}\Big(\frac{2\pi n}{T}\Big)^{-1}\Big[2\pi\mathbf{I}_{\mathbf{Y}\hat{\psi}(\mathbf{Y})}\Big(\frac{2\pi n}{T}\Big)-\mathbf{I}\Big]\Big\}. \tag{8}$$

where

$$\mathbf{I}_{\mathbf{Y}\hat{\psi}(\mathbf{Y})}\Big(\frac{2\pi n}{T}\Big) = \frac{1}{2\pi T}\mathbf{d}_{\mathbf{Y}}\Big(\frac{2\pi n}{T}\Big)\mathbf{d}^{\mathrm{T}}_{\hat{\psi}(\mathbf{Y})}\Big(2\pi\frac{T-n}{T}\Big) \tag{9}$$

is the cross periodogram between $\{\mathbf{Y}(t)\}$ and $\{\hat{\psi}[\mathbf{Y}(t)]\}$.

Setting the gradient to 0 one gets the estimating equations and in the limit ($T\to\infty$) the theoretical estimating equations

$$\int_0^{2\pi}\mathrm{tr}\Big\{\frac{\partial\mathbf{B}(\omega)}{\partial\theta_\mu}\mathbf{B}(\omega)^{-1}[2\pi f_{\mathbf{Y},\psi(\mathbf{Y})}(\omega)-\mathbf{I}]\Big\}\frac{d\omega}{2\pi} = \mathbf{0}$$

where $f_{\mathbf{Y}\psi(\mathbf{Y})}(\omega) = (2\pi)^{-1}\sum_{u=-\infty}^{\infty}\mathrm{E}\{\mathbf{Y}(u)\psi[\mathbf{Y}(0)]^{\mathrm{T}}\}e^{-iu\omega}$ is the cross spectral density between the processes $\{\mathbf{Y}(t)\}$ and $\{\psi[\mathbf{Y}(t)]\}$. This equation is satisfied if $\{\mathbf{B}(u)\}$ is a separating-deconvoluting filter, that is if it yields sequences $\{Y_k(u)\}$ which are temporally independent and mutually independent. Indeed, in this case, the spectral density $f_{\mathbf{Y}\psi(\mathbf{Y})}$ is a constant diagonal matrix, which equals $\mathbf{I}/2\pi$ since $\mathrm{E}[Y_k\psi_k(Y_k)] = 1$ (see for ex. [1]).

3. THE QUASI-NEWTON ALGORITHM

One may solve the estimating equation by the quasi-Newton algorithm. This algorithm computes the solution of the equation $l(\theta) = 0$ through the iteration $\theta^{(\nu)}\mapsto\theta^{(\nu+1)} =$ the solution of the equation $K(\theta^{(\nu)})(\theta^{(\nu+1)}-\theta^{(\nu)}) = -l(\theta^{(\nu)})$, where $K(\theta)$ is some approximation of the Jacobian $\partial l/\partial\theta$ of the map $\theta\mapsto l(\theta)$.

In the present case, $l(\theta)$ is given by (7) or (8). To construct $K(\theta)$, we shall treat the sequences $\{Y_1(t)\},\ldots,\{Y_K(t)\}$ as temporally independent and independent among themselves, which is justified (for most value of t which are not near 0 or $T-1$), if θ is near the true parameter and T is large enough.

The derivative of the first term in (7) can be splitted into $S_1+S_2+S_3$ where

$$S_1 = \mathrm{tr}\Big\{\sum_u\frac{\partial^2\mathbf{B}(u)}{\partial\theta_\mu\partial\theta_\nu}\frac{1}{T}\sum_{t=0}^{T-1}\mathbf{X}(t-u \bmod T)\hat{\psi}^{\mathrm{T}}[\mathbf{Y}(t)]\Big\},$$

$$S_2 = \mathrm{tr}\Big\{\frac{1}{T}\sum_{t=0}^{T-1}\frac{\partial\mathbf{Y}(t)}{\partial\theta_\mu}\frac{\partial\mathbf{Y}^{\mathrm{T}}(t)}{\partial\theta_\nu}\mathrm{diag}\{\hat{\psi}'[\mathbf{Y}(t)]\}\Big\}$$

$'$ denoting the derivative and $\mathrm{diag}\{\hat{\psi}'[\mathbf{Y}(t)]\}$ denoting the diagonal matrix with diagonal elements $\hat{\psi}_1'[Y_1(t)]$, $\ldots$, $\hat{\psi}_K'[Y_K(t)]$, and

$$S_3 = \mathrm{tr}\Big\{\frac{1}{T}\sum_{t=0}^{T-1}\frac{\partial\mathbf{Y}(t)}{\partial\theta_\mu}\frac{\partial\hat{\psi}^{\mathrm{T}}}{\partial\theta_\nu}[\mathbf{Y}(t)]\Big\}$$

Consider first S_1. Since $\partial^2\mathbf{B}(u)/\partial\theta_\mu\partial\theta_\nu$ should converge to zero quickly as $u\to\pm\infty$, one may restrict oneself to small u. Then for large T one may approximate $T^{-1}\sum_{t=0}^{T-1}\mathbf{X}(t-u \bmod T)\hat{\psi}^{\mathrm{T}}[\mathbf{Y}(t)]$ by $\mathrm{E}\{\mathbf{X}(-u)\psi^{\mathrm{T}}[\mathbf{Y}(0)]\} = \int_0^{2\pi}\mathbf{B}(\omega)^{-1}f_{\mathbf{Y}\psi(\mathbf{Y})}(\omega)e^{-i\omega u}d\omega$. Thus S_1 equals approximately

$$\int_0^{2\pi}\mathrm{tr}\Big[\frac{\partial^2\mathbf{B}(\omega)}{\partial\theta_\mu\partial\theta_\nu}\mathbf{B}(\omega)^{-1}f_{\mathbf{Y}\psi(\mathbf{Y})}(\omega)\Big]d\omega.$$

By treating $\{Y_1(t)\},\ldots,\{Y_K(t)\}$ as temporally independent and independent among themselves, $f_{\mathbf{Y}\psi(\mathbf{Y})}(\omega)$ reduces to $\mathbf{I}/(2\pi)$ Hence, S_1 minus the derivative of the last term in (7) can be approximated by

$$\int_0^{2\pi}\mathrm{tr}\Big[\frac{\partial\mathbf{B}(\omega)}{\partial\theta_\mu}\mathbf{B}(\omega)^{-1}\frac{\partial\mathbf{B}(\omega)}{\partial\theta_\nu}\mathbf{B}(\omega)^{-1}\Big]\frac{d\omega}{2\pi}.$$

Consider now S_2. Let $\{\mathbf{B}^\dagger(u)\}$ be the inverse sequence (in the convolution sense) of $\{\mathbf{B}(u)\}$, then by (4) $\mathbf{X}(t \bmod T) = \sum_u\mathbf{B}^\dagger(u)\mathbf{Y}(t-u)$. Hence by (6) $\partial\mathbf{Y}(t)/\partial\theta_\mu = \sum_u\mathbf{C}_\mu(u)\mathbf{Y}(t-u)$ where $\mathbf{C}_\mu(u) = \sum_v[\partial\mathbf{B}(v)/\partial\theta_\mu]\mathbf{B}^\dagger(u-v)$. Thus S_2 can be written as the trace of

$$\sum_u\sum_v\frac{1}{T}\sum_{t=0}^{T-1}\mathbf{C}_\mu(u)\mathbf{Y}(t-u)\mathbf{Y}^{\mathrm{T}}(t-v)\mathbf{C}_\nu^{\mathrm{T}}(v)\mathrm{diag}\{\hat{\psi}'[\mathbf{Y}(t)]\},$$

As before, on may restrict oneself to small u and v. Then for large T, one may approximate $T^{-1}\sum_{t=1}^{T-1} Y_j(t-u)Y_l(t-v)\psi_k'[Y_k(t)]$ by 0 if $(u,j)\neq(v,l)$, by $\mathrm{E}Y_j^2\mathrm{E}\psi_k'(Y)$ if $(u,j)=(v,l)\neq(0,k)$, and by $\mathrm{E}Y_k^2\mathrm{E}\psi_k'(Y)+\mathrm{cov}[Y_1^2\psi_k'(Y_k)]$ if $u=v=0$ and $j=l=k$. With these approximations, a somewhat tedious algebra yields

$$S_2 \approx \sum_{j,k=1}^{K}\sum_u \mathrm{E}Y_j^2\mathrm{E}\psi_k'(Y_k)C_{\mu,kj}(u)C_{\nu,kj}(u) + \sum_{k=1}^{K} C_{\mu,kk}(0)C_{\nu,kk}(0)\,\mathrm{cov}[Y_k^2,\psi'(Y_k)],$$

$C_{\mu,jk}(u)$ denoting the general element of $\mathbf{C}_\mu(u)$. By the Parseval equality, one may rewrite the first term in the above right hand side as

$$\int_0^{2\pi}\sum_{j,k=1}^{K}\mathrm{E}Y_j^2\mathrm{E}\psi_k'(Y_k)\times \Big[\frac{\partial\mathbf{B}(\omega)}{\partial\theta_\mu}\mathbf{B}(\omega)^{-1}\Big]_{kj}\Big[\frac{\partial\mathbf{B}(-\omega)}{\partial\theta_\nu}\mathbf{B}(-\omega)^{-1}\Big]_{kj}\frac{d\omega}{2\pi}.$$

Consider finally S_3. Since the function $\partial\hat\psi_k/\partial\theta_\nu$ does not depend on the time index, one may approximate the time average $T^{-1}\sum_{t=0}^{T-1}\mathbf{Y}(t-u)(\partial\hat\psi/\partial\theta_\nu)^{\mathrm{T}}[\mathbf{Y}(t)]$ by 0 if $u\neq 0$. Further, since this function depends only on $Y_k(0),\ldots,Y_k(T-1)$ which are independent of the $Y_j(t)$, for $j\neq k$, we can also approximate the time average $T^{-1}\sum_{t=0}^{T-1}Y_j(t)(\partial\hat\psi/\partial\theta_\nu)[Y_k(t)]$ by 0. Hence

$$S_3 \approx \sum_{k=1}^{K} C_{\mu,kk}(0)\frac{1}{T}\sum_{t=0}^{T-1}Y_k(t)\frac{\partial\hat\psi_k}{\partial\theta_\nu}[Y_k(t)].$$

But our estimator $\hat\psi_k$ satisfies $T^{-1}\sum_{t=0}^{T-1}Y_k(t)\hat\psi_k[Y_k(t)]=1$ (see [6], eq. 9), therefore

$$\frac{1}{T}\sum_{t=0}^{T-1}Y_k(t)\frac{\partial\hat\psi_k}{\partial\theta_\nu}[Y_k(t)] = -\frac{1}{T}\sum_{t=0}^{T-1}\Big\{\hat\psi[Y_k(t)]+Y_k(t)\hat\psi_k'[Y_k(t)]\Big\}\frac{\partial Y_k(t)}{\partial\theta_\nu}$$

By a similar calculation as before, the last right hand side may be approximated by $-C_{\nu,kk}(0)\{1+\mathrm{E}[Y_k^2\hat\psi_k'(Y_k)]\}$. Thus, one gets finally

$$S_3 \approx -\sum_{k=1}^{K} C_{\mu,kk}(0)C_{\nu,kk}(0)\{1+\mathrm{E}[Y_k^2\psi_k'(Y_k)]\}.$$

Combining the above results and using now the notation $\overline{\partial\mathbf{B}/\partial\theta_\mu\mathbf{B}^{-1}}$ for $\mathbf{C}_\mu(0)$ as it equals the average $\int_0^{2\pi}\partial\mathbf{B}(\omega)/\partial\theta_\mu\mathbf{B}(\omega)^{-1}d\omega/(2\pi)$ of the $\partial\mathbf{B}(\omega)/\partial\theta_\mu\mathbf{B}(\omega)^{-1}$ over $[0,2\pi]$, one gets

$$K_{\mu\nu}(\theta)\approx\sum_{j,k=1}^{K}\int_0^{2\pi}\Big\{\mathrm{E}\psi_k'(Y_k)\mathrm{E}Y_j^2 \times\Big[\frac{\partial\mathbf{B}(\omega)}{\partial\theta_\mu}\mathbf{B}(\omega)^{-1}\Big]_{kj}\Big[\frac{\partial\mathbf{B}(-\omega)}{\partial\theta_\nu}\mathbf{B}(-\omega)^{-1}\Big]_{kj} + \Big[\frac{\partial\mathbf{B}(\omega)}{\partial\theta_\mu}\mathbf{B}(\omega)^{-1}\Big]_{kj}\Big[\frac{\partial\mathbf{B}(\omega)}{\partial\theta_\nu}\mathbf{B}(\omega)^{-1}\Big]_{jk}\Big\}\frac{d\omega}{2\pi} - \sum_{k=1}^{K}[\mathrm{E}\psi_k'(Y_k)\mathrm{E}Y_k^2+1]\Big(\overline{\frac{\partial\mathbf{B}}{\partial\theta_\mu}\mathbf{B}^{-1}}\Big)_{kk}\Big(\overline{\frac{\partial\mathbf{B}}{\partial\theta_\nu}\mathbf{B}^{-1}}\Big)_{kk}.$$

4. THE PURE DECONVOLUTION CASE

This case corresponds to $K=1$, hence $\mathbf{X}(t)$, $\mathbf{Y}(t)$ and $\mathbf{B}$ are all scalar and will be denoted as $X(t)$, $Y(t)$ and B and we drop the index k in ψ_k.

The expression (8) for the gradient in this case reduces to

$$\frac{1}{T}\sum_{n=0}^{T-1}\frac{\partial\log B(2\pi n/T)}{\partial\theta_\mu}\Big[2\pi I_{Y\hat\psi(Y)}\Big(\frac{2\pi n}{T}\Big)-1\Big], \qquad (10)$$

where $I_{Y\hat\psi(Y)}(2\pi n/T)$ is defined similarly to (9), except that it is now a scalar.

The approximate Hessian can also be written more compactly, putting $\kappa=\mathrm{E}Y^2\mathrm{E}\psi'(Y)$,

$$K_{\mu\nu}(\theta)\approx\int_0^{2\pi}\Big[\kappa\frac{\partial\log B(-\omega)}{\partial\theta_\mu}+\frac{\partial\log B(\omega)}{\partial\theta_\nu}\Big]\frac{\partial\log B(\omega)}{\partial\theta_\mu}\frac{d\omega}{2\pi} - \sum_{k=1}^{K}\{\kappa+1\}\overline{\frac{\partial\log B}{\partial\theta_\mu}}\,\overline{\frac{\partial\log B}{\partial\theta_\nu}}.$$

The above formula shows that it is of interest to parameterize $\log B(\omega)$ instead of $B(\omega)$. Further, by separating the real and imaginary part of $\partial\log B(\omega)/\partial\theta_\mu$, one can rewrite $K_{\mu\nu}(\theta)$ as,

$$K_{\mu\nu}(\theta)\approx\int_0^{2\pi}\Big\{(\kappa+1)\Big[\Re\frac{\partial\log B(\omega)}{\partial\theta_\mu}-\overline{\frac{\partial\log B}{\partial\theta_\mu}}\Big]\Big[\Re\frac{\partial\log B(\omega)}{\partial\theta_\nu}-\overline{\frac{\partial\log B}{\partial\theta_\nu}}\Big] + (\kappa-1)\Big[\Im\frac{\partial\log B(\omega)}{\partial\theta_\mu}\Im\frac{\partial\log B(\omega)}{\partial\theta_\nu}\Big\}\frac{d\omega}{2\pi},$$

$\Re$ and $\Im$ denoting the real and imaginary parts. This formula shows that it is of interest to parameterize the real and imaginary parts of $\log B(\omega)$ by two different sets of parameters, since then there is a *decoupling* between these sets in the quasi-Newton algorithm.

An simple interesting parameterization for $\log B(\omega)$ is

$$\log B(\omega)=\theta_0+\sum_{\mu=1}^{m}[\theta_\mu\cos(\mu\omega)+i\theta_{\mu+m}\sin(\mu\omega)].$$

The parameters $\theta_0,\ldots,\theta_m$ and $\theta_{m+1},\ldots,\theta_{2m}$ specify the real and imaginary parts of $\log\mathbf{B}(\omega)$ respectively (these parts are even and odd functions respectively, hence the use of the cosine and sine functions to represent them). The parameter θ_0 correspond to the scale of the sources and cannot be estimated. We may (and will) put it to 0. The expression (10) for the gradient reduces to $(r_\mu+r_{-\mu})/2$, for $\mu=1,\ldots,m$ and r_m-r_{-m} for $\mu=m+1,\ldots,2m$, where

$$r_\mu=\frac{2\pi}{T}\sum_{n=0}^{T-1}e^{i2\pi n\mu/T}I_{Y,\hat\psi(Y)}\Big(\frac{2\pi n}{T}\Big)=\frac{1}{T}\sum_{n=0}^{T-1}Y(t+\mu)\psi[Y(t)] \qquad (11)$$

are the circular cross covariances between $\{Y(t)\}$ and $\{\psi[Y(t)]\}$. Thus, the quasi-Newton algorithm reduces to the fixed point iteration

$$\theta_\mu \mapsto \theta_\mu - \begin{cases} (r_\mu + r_{-\mu}/(\hat\kappa+1), & \mu = 1,\dots,m \\ (r_{\mu-m} - r_{m-\mu})/(\hat\kappa-1), & \mu = m+1,\dots,2m \end{cases},$$

$\hat\kappa$ denoting the current estimate for κ. It is also of interest to work with the parameters $\vartheta_\mu = (\theta_\mu + \theta_{\mu+m})/2$ and $\vartheta_{-\mu} = (\theta_\mu - \theta_{\mu+m})/2$, $\mu = 1,\dots,m$ so that one has

$$\log B(\omega) = \sum_{\mu=-m}^{m} \vartheta_\mu e^{i\mu\omega}, \qquad (\vartheta_0 = \theta_0).$$

The quasi-Newton algorithm for these parameters is then

$$\vartheta_\mu \mapsto \vartheta_\mu - \frac{\kappa r_\mu - r_{-\mu}}{\kappa^2 - 1}, \qquad 0 < |\mu| \le m. \tag{12}$$

Note that the criterion (5) reduces to simply $\hat H(Y)$, since $\int \log B(\omega) d\omega/(2\pi) = \vartheta_0 = 0$. This criterion can be computed to ensure that it is decreased at each iteration, otherwise the Newton step could be reduced by a factor < 1 so that it is so.

5. A SIMULATION EXAMPLE

We consider the observation model $X(t) = \rho' S(t-1) + S(t) + \rho'' S(t+1)$ where $\{S(t)\}$ is the source sequence. Thus $B(\omega)$ is proportional to $1/(1+\rho' e^{-i\omega} + \rho'' e^{i\omega})$. It can be shown that for $|\rho' + \rho''| < 1$, one can factorize $(1 + \rho' e^{-i\omega} + \rho'' e^{i\omega})$ as $c(1 + z' e^{-i\omega})(1 + z'' e^{i\omega})$, with $c = 1/2 + \sqrt{1/4 - \rho'\rho''}$ and $z' = \rho'/c$, $z'' = \rho''/c$ and both $|z'|$ and $|z''|$ are strictly less than 1. Thus one has the expansion

$$\log B(\omega) = \sum_{j=-\infty}^{-1} \frac{(-1)^j z'^{|j|}}{|j|} e^{-ij\omega} + \sum_{j=1}^{\infty} \frac{(-1)^j z''^j}{j} e^{ij\omega} + \text{Const.}$$

j	1	2	3	4	5
$(z''^j + z'^j)/j$	1.1117	.3101	.1157	.0487	.0220
$(z''^j - z'^j)/j$	0.0654	.0364	.0202	.0113	.0063
j	6	7	8	9	10
$(z''^j + z'^j)/j$	.0103	.0050	.0025	.0013	.0007
$(z''^j - z'^j)/j$	.0035	.0020	.0011	.0006	.0003

Table 1: Value of $(z''^j \pm z'^j)/j$ for $j = 1,\dots,10$ ($\rho' = 0.4$, $\rho'' = 0.45$)

We take $\rho' = 0.4$, $\rho'' = 0.45$, which yields the values of $(z''^j \pm z'^j)/j$ reported in the table 1. One can see from this table that one may truncate without incurring much error the expansion for $\log \mathbf{B}(\omega)$ at index $m = \pm 10$ and thus consider the value reported in table 1 as the "true value" of $|\theta_j|$ (first row) and $|\theta_{m+j}|$ (second row). We generate 1000 sequences of observations of length $T = 512$, using as source distribution the bilateral exponential distribution. For each sequence of observations, we apply our algorithm with $m = 10$ and the entropy $H(Y)$ and score function ψ_Y are estimated by the method of [6]. The mean and standard deviation of the resulting estimators $\hat\theta_j$ are reported in table 2.

j	$\hat\theta_j$		$\hat\theta_{10+j}$	
	mean	std. dev	mean	std. dev
1	-1.1029	0.0417	-0.0613	0.0954
2	0.3098	0.0420	0.0300	0.0957
3	-0.1099	0.0408	-0.0154	0.0950
4	0.0494	0.0431	0.0085	0.0959
5	-0.0207	0.0405	-0.0016	0.0923
6	0.0119	0.0424	-0.0002	0.0935
7	-0.0016	0.0429	-0.0025	0.0950
8	0.0022	0.0418	-0.0024	0.0942
9	0.0020	0.0416	0.0007	0.0984
10	0.0033	0.0428	0.0008	0.0901

Table 2: Simulation results for the estimator $\hat\theta_j$ based on 1000 repetitions; sample size $T = 512$.

The algorithm converges quite fast: the mean number of iteration is 4.2. Note that the gain of the separating filter $|B(\omega)|^2$ must be proportional to the spectral density of the observed sequence $\{X(t)\}$, which admits $|d_X(2\pi n/T)|^2/(2\pi T)$ as a raw estimate. Therefore the parameters $\theta_1,\dots,\theta_m$ specifying the real part of $\log B(\omega)$ can be estimated directly from the $|d_X(2\pi n/T)|^2$ and such estimates actually serve to initialize our algorithm.

6. CONCLUSION

We have provided a solution of the blind separation-deconvolution problem, based on the mutual information criterion and obtained formulas for the quasi Newton algorithm. Formula for the asymptotic covariance matrix of the estimator can be obtained as well but not presented due to lack of space. In the pure deconvolution case, we propose a parameterization which leads to a very simple quasi Newton algorithm. The algorithm has been validated by simulation. We haven't provided simulations for the quasi Newton algorithm in the general separation-deconvolution case because of lack of space and also because it is quite complex. We are currently developed a simplified version, to be reported later.

REFERENCES

[1] D. T. Pham, "Blind separation of instantaneous mixture of sources via an independent component analysis," *IEEE Transactions on Signal Processing*, vol. 44, no. 11, pp. 2768–2779, Nov. 1996.

[2] T. Cover and J. A. Thomas, *Elements of Information Theory*. New-York: Wiley, 1991.

[3] D. T. Pham, "Mutual information approach to blind separation of stationary sourcs," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1935–1946, July 2002.

[4] A. Taleb, J. S. i Casal, and C. Jutten, "Quasi-nonparametric blind inversion of Wiener systems," *IEEE Trans. on Signal Processing*, vol. 49, no. 5, pp. 917–924, 2001.

[5] D. L. Donoho, *On minimum entropy deconvolution*. New-York: Academic Press, 1981, pp. 565–609.

[6] D. T. Pham, "Fast algorithms for mutual information based independent component analysis," *IEEE Trans. on Signal Processing*, vol. 52, no. 10, pp. 2690–2700, 2004.